

8

Semiconductor Crystals

BAND GAP	187
EQUATIONS OF MOTION	191
Physical derivation of $\hbar\mathbf{k} = \mathbf{F}$	193
Holes	194
Effective mass	197
Physical interpretation of the effective mass	198
Effective masses in semiconductors	200
Silicon and germanium	202
INTRINSIC CARRIER CONCENTRATION	205
Intrinsic mobility	208
IMPURITY CONDUCTIVITY	209
Donor states	209
Acceptor states	211
Thermal ionization of donors and acceptors	213
THERMOELECTRIC EFFECTS	214
SEMIMETALS	215
SUPERLATTICES	216
Bloch oscillator	217
Zener tunneling	217
SUMMARY	217
PROBLEMS	218
1. Impurity orbits	218
2. Ionization of donors	218

NOTE: The discussion of carrier orbits in applied fields is continued in Chapter 9.

- 3. Hall effect with two carrier types 218
- 4. Cyclotron resonance for a spheroidal energy surface 219
- 5. Magnetoresistance with two carrier types 219

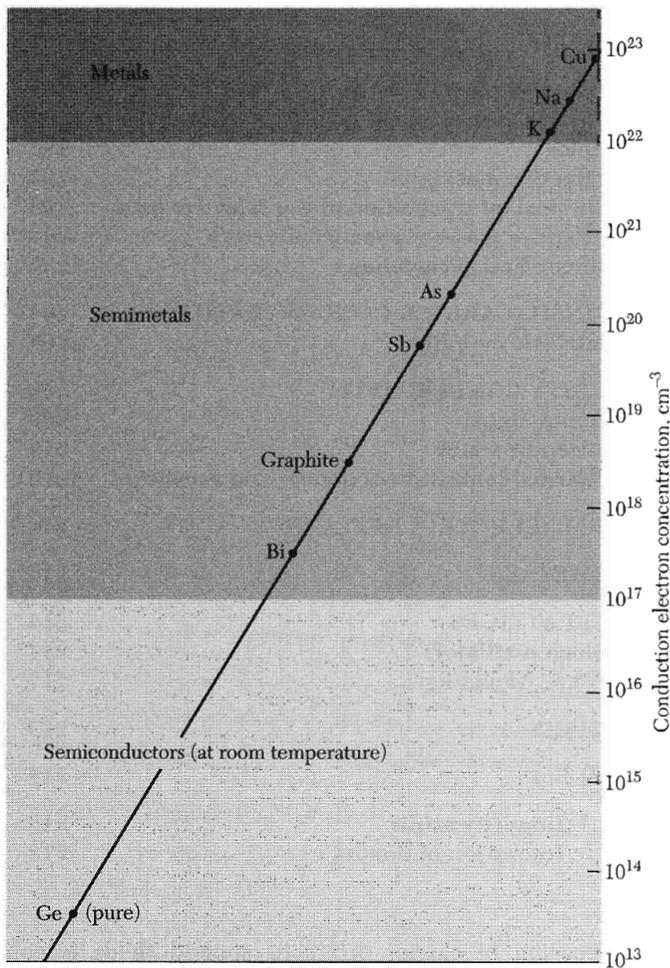


Figure 1 Carrier concentrations for metals, semimetals, and semiconductors. The semiconductor range may be extended upward by increasing the impurity concentration, and the range can be extended downward to merge eventually with the insulator range.

CHAPTER 8: SEMICONDUCTOR CRYSTALS

Carrier concentrations representative of metals, semimetals, and semiconductors are shown in Fig. 1. Semiconductors are generally classified by their electrical resistivity at room temperature, with values in the range of 10^{-2} to 10^9 ohm-cm, and strongly dependent on temperature. At absolute zero a pure, perfect crystal of most semiconductors will be an insulator, if we arbitrarily define an insulator as having a resistivity above 10^{14} ohm-cm.

Devices based on semiconductors include transistors, switches, diodes, photovoltaic cells, detectors, and thermistors. These may be used as single circuit elements or as components of integrated circuits. We discuss in this chapter the central physical features of the classical semiconductor crystals, particularly silicon, germanium, and gallium arsenide.

Some useful nomenclature: the semiconductor compounds of chemical formula AB , where A is a trivalent element and B is a pentavalent element, are called III-V (three-five) compounds. Examples are indium antimonide and gallium arsenide. Where A is divalent and B is hexavalent, the compound is called a II-VI compound; examples are zinc sulfide and cadmium sulfide. Silicon and germanium are sometimes called diamond-type semiconductors, because they have the crystal structure of diamond. Diamond itself is more an insulator rather than a semiconductor. Silicon carbide SiC is a IV-IV compound.

A highly purified semiconductor exhibits intrinsic conductivity, as distinguished from the impurity conductivity of less pure specimens. In the **intrinsic temperature range** the electrical properties of a semiconductor are not essentially modified by impurities in the crystal. An electronic band scheme leading to intrinsic conductivity is indicated in Fig. 2. The conduction band is vacant at absolute zero and is separated by an energy gap E_g from the filled valence band.

The **band gap** is the difference in energy between the lowest point of the conduction band and the highest point of the valence band. The lowest point in the conduction band is called the **conduction band edge**; the highest point in the valence band is called the **valence band edge**.

As the temperature is increased, electrons are thermally excited from the valence band to the conduction band (Fig. 3). Both the electrons in the conduction band and the vacant orbitals or holes left behind in the valence band contribute to the electrical conductivity.

BAND GAP

The intrinsic conductivity and intrinsic carrier concentrations are largely controlled by $E_g/k_B T$, the ratio of the band gap to the temperature. When this ratio is large, the concentration of intrinsic carriers will be low and the

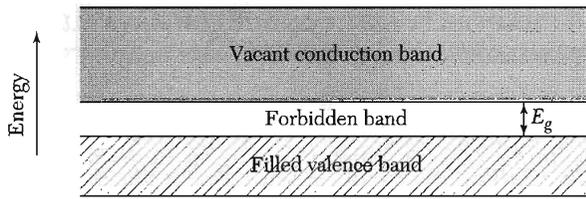


Figure 2 Band scheme for intrinsic conductivity in a semiconductor. At 0 K the conductivity is zero because all states in the valence band are filled and all states in the conduction band are vacant. As the temperature is increased, electrons are thermally excited from the valence band to the conduction band, where they become mobile. Such carriers are called “intrinsic.”

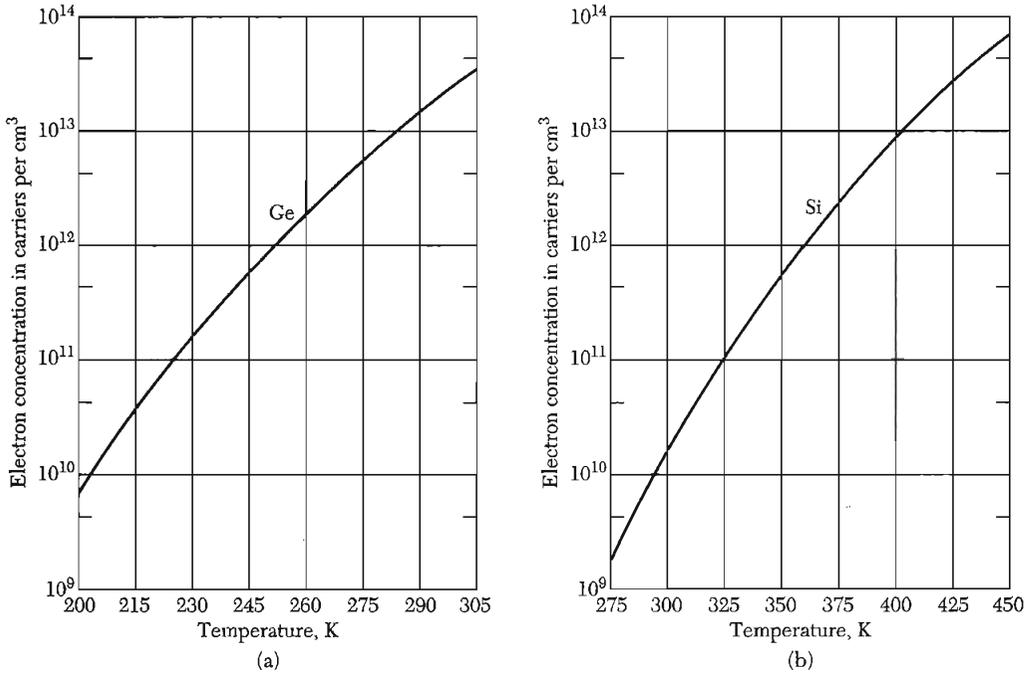


Figure 3 Intrinsic electron concentration as a function of temperature for (a) germanium and (b) silicon. Under intrinsic conditions the hole concentration is equal to the electron concentration. The intrinsic concentration at a given temperature is higher in Ge than in Si because the energy gap is narrower in Ge (0.66 eV) than in Si (1.11 eV). (After W. C. Dunlap.)

conductivity will be low. Band gaps of representative semiconductors are given in Table 1. The best values of the band gap are obtained by optical absorption.

In a **direct absorption process** the threshold of continuous optical absorption at frequency ω_g measures the band gap $E_g = \hbar\omega_g$ as shown in Figs. 4a and 5a. A photon is absorbed by the crystal with the creation of an electron and a hole.

In the **indirect absorption process** in Figs. 4b and 5b the minimum energy gap of the band structure involves electrons and holes separated by a

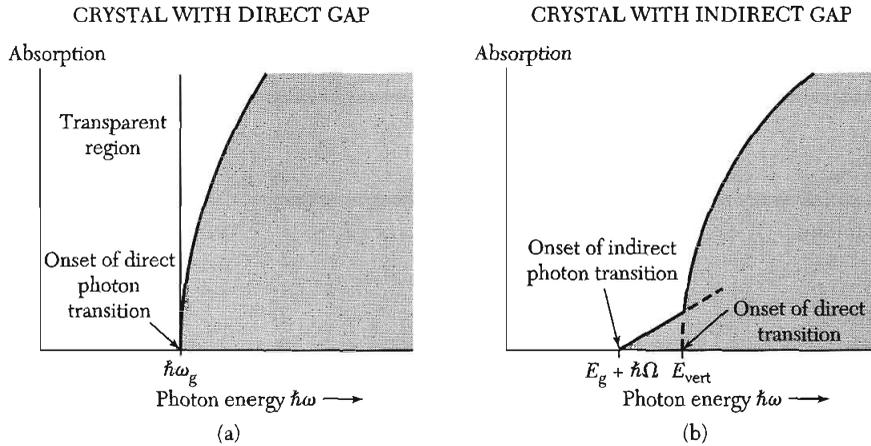


Figure 4 Optical absorption in pure insulators at absolute zero. In (a) the threshold determines the energy gap as $E_g = \hbar\omega_g$. In (b) the optical absorption is weaker near the threshold: at $\hbar\omega = E_g + \hbar\Omega$ a photon is absorbed with the creation of three particles: a free electron, a free hole, and a phonon of energy $\hbar\Omega$. In (b) the energy E_{vert} marks the threshold for the creation of a free electron and a free hole, with no phonon involved. Such a transition is called vertical; it is similar to the direct transition in (a). These plots do not show absorption lines that sometimes are seen lying just to the low energy side of the threshold. Such lines are due to the creation of a bound electron-hole pair, called an exciton.

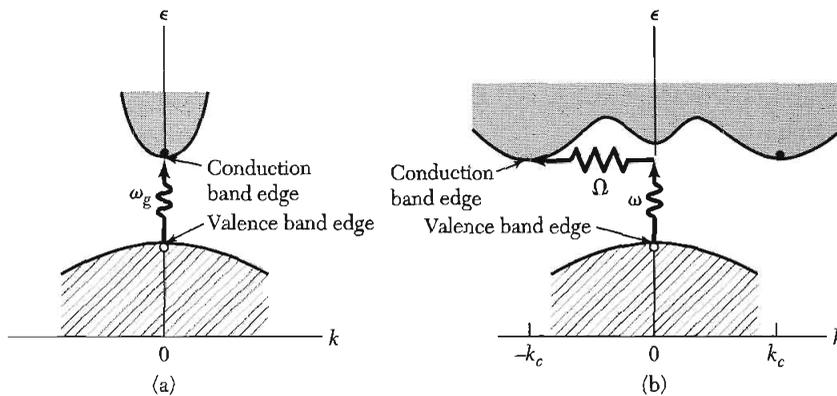


Figure 5 In (a) the lowest point of the conduction band occurs at the same value of \mathbf{k} as the highest point of the valence band. A direct optical transition is drawn vertically with no significant change of \mathbf{k} , because the absorbed photon has a very small wavevector. The threshold frequency ω_g for absorption by the direct transition determines the energy gap $E_g = \hbar\omega_g$. The indirect transition in (b) involves both a photon and a phonon because the band edges of the conduction and valence bands are widely separated in \mathbf{k} space. The threshold energy for the indirect process in (b) is greater than the true band gap. The absorption threshold for the indirect transition between the band edges is at $\hbar\omega = E_g + \hbar\Omega$, where Ω is the frequency of an emitted phonon of wavevector $\mathbf{K} \equiv -\mathbf{k}_g$. At higher temperatures phonons are already present; if a phonon is absorbed along with a photon, the threshold energy is $\hbar\omega = E_g - \hbar\Omega$. Note: The figure shows only the threshold transitions. Transitions occur generally between almost all points of the two bands for which the wavevectors and energy can be conserved.

Table 1 Energy gap between the valence and conduction bands
(*i* = indirect gap; *d* = direct gap)

Crystal	Gap	E_g , eV		Crystal	Gap	E_g , eV	
		0 K	300 K			0 K	300 K
Diamond	<i>i</i>	5.4		SiC(hex)	<i>i</i>	3.0	—
Si	<i>i</i>	1.17	1.11	Tc	<i>d</i>	0.33	—
Ge	<i>i</i>	0.744	0.66	HgTe ^a	<i>d</i>	-0.30	
α Sn	<i>d</i>	0.00	0.00	PbS	<i>d</i>	0.286	0.34–0.37
InSb	<i>d</i>	0.23	0.17	PbSe	<i>i</i>	0.165	0.27
InAs	<i>d</i>	0.43	0.36	PbTe	<i>i</i>	0.190	0.29
InP	<i>d</i>	1.42	1.27	CdS	<i>d</i>	2.582	2.42
GaP	<i>i</i>	2.32	2.25	CdSe	<i>d</i>	1.840	1.74
GaAs	<i>d</i>	1.52	1.43	CdTe	<i>d</i>	1.607	1.44
GaSb	<i>d</i>	0.81	0.68	SnTe	<i>d</i>	0.3	0.18
AlSb	<i>i</i>	1.65	1.6	Cu ₂ O	<i>d</i>	2.172	—

^aHgTe is a semimetal; the bands overlap.

substantial wavevector \mathbf{k}_c . Here a direct photon transition at the energy of the minimum gap cannot satisfy the requirement of conservation of wavevector, because photon wavevectors are negligible at the energy range of interest. But if a phonon of wavevector \mathbf{K} and frequency Ω is created in the process, then we can have

$$\mathbf{k}(\text{photon}) = \mathbf{k}_c + \mathbf{K} \cong \mathbf{0} ; \quad \hbar\omega = E_g + \hbar\Omega ,$$

as required by the conservation laws. The phonon energy $\hbar\Omega$ will generally be much less than E_g ; a phonon even of high wavevector is an easily accessible source of crystal momentum because the phonon energies are characteristically small (~ 0.01 to 0.03 eV) in comparison with the energy gap. If the temperature is high enough that the necessary phonon is already thermally excited in the crystal, it is possible also to have a photon absorption process in which the phonon is absorbed.

The band gap may also be deduced from the temperature dependence of the conductivity or of the carrier concentration in the intrinsic range. The carrier concentration is obtained from measurements of the Hall voltage (Chapter 6), sometimes supplemented by conductivity measurements. Optical measurements determine whether the gap is direct or indirect. The band edges in Ge and in Si are connected by indirect transitions; the band edges in InSb and GaAs are connected by a direct transition (Fig. 6). The gap in α Sn is

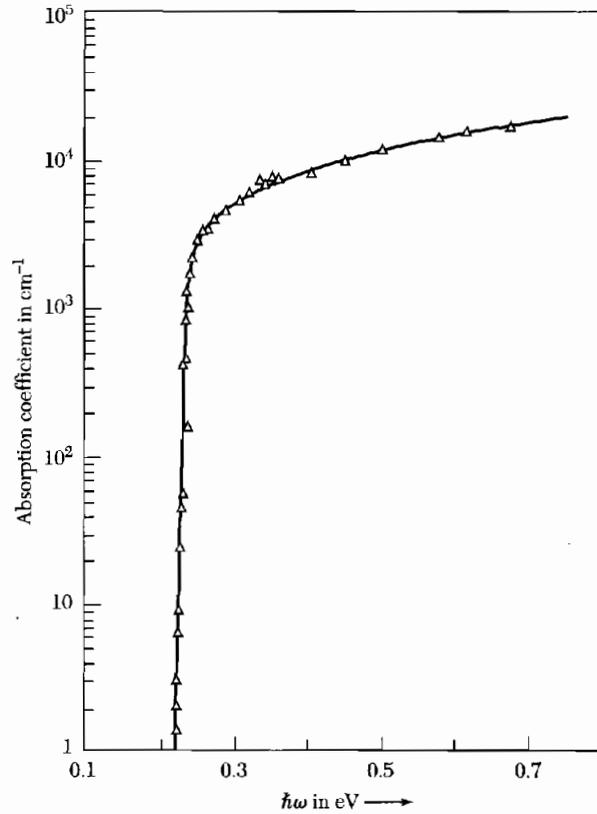


Figure 6 Optical absorption in pure indium antimonide, InSb. The transition is direct because both conduction and valence band edges are at the center of the Brillouin zone, $\mathbf{k} = 0$. Notice the sharp threshold. (After G. W. Gobeli and H. Y. Fan.)

direct and is exactly zero; HgTe and HgSe are semimetals and have negative gaps—the conduction and valence bands overlap.

EQUATIONS OF MOTION

We derive the equation of motion of an electron in an energy band. We look at the motion of a wave packet in an applied electric field. Suppose that the wave packet is made up of wavefunctions assembled near a particular wavevector k . The group velocity by definition is $v_g = d\omega/dk$. The frequency associated with a wavefunction of energy ϵ by quantum theory is $\omega = \epsilon/\hbar$, and so

$$v_g = \hbar^{-1} d\epsilon/dk \quad \text{or} \quad \mathbf{v} = \hbar^{-1} \nabla_{\mathbf{k}} \epsilon(\mathbf{k}) . \quad (1)$$

The effects of the crystal on the electron motion are contained in the dispersion relation $\epsilon(\mathbf{k})$.

The work $\delta\epsilon$ done on the electron by the electric field E in the time interval δt is

$$\delta\epsilon = -eEv_g \delta t \quad (2)$$

We observe that

$$\delta\epsilon = (d\epsilon/dk)\delta k = \hbar v_g \delta k \quad (3)$$

using (1). On comparing (2) with (3) we have

$$\delta k = -(eE/\hbar)\delta t \quad (4)$$

whence $\hbar dk/dt = -eE$.

We may write (4) in terms of the external force \mathbf{F} as

$$\hbar \frac{d\mathbf{k}}{dt} = \mathbf{F} \quad (5)$$

This is an important relation: in a crystal $\hbar d\mathbf{k}/dt$ is equal to the external force on the electron. In free space $d(m\mathbf{v})/dt$ is equal to the force. We have not overthrown Newton's second law of motion: the electron in the crystal is subject to forces from the crystal lattice as well as from external sources.

The force term in (5) also includes the electric field and the Lorentz force on an electron in a magnetic field, under ordinary conditions where the magnetic field is not so strong that it breaks down the band structure. Thus the equation of motion of an electron of group velocity \mathbf{v} in a constant magnetic field \mathbf{B} is

$$\text{(CGS)} \quad \hbar \frac{d\mathbf{k}}{dt} = -\frac{e}{c} \mathbf{v} \times \mathbf{B} ; \quad \text{(SI)} \quad \hbar \frac{d\mathbf{k}}{dt} = -e\mathbf{v} \times \mathbf{B} \quad (6)$$

where the right-hand side of each equation is the Lorentz force on the electron. With the group velocity $\mathbf{v} = \hbar^{-1} \text{grad}_{\mathbf{k}}\epsilon$, the rate of change of the wavevector is

$$\text{(CGS)} \quad \frac{d\mathbf{k}}{dt} = -\frac{e}{\hbar^2 c} \nabla_{\mathbf{k}}\epsilon \times \mathbf{B} ; \quad \text{(SI)} \quad \frac{d\mathbf{k}}{dt} = -\frac{e}{\hbar^2} \nabla_{\mathbf{k}}\epsilon \times \mathbf{B} \quad (7)$$

where now both sides of the equation refer to the coordinates in \mathbf{k} space.

We see from the vector cross-product in (7) that in a magnetic field an electron moves in \mathbf{k} space in a direction normal to the direction of the gradient of the energy ϵ , so that **the electron moves on a surface of constant energy**. The value of the projection $k_{\mathbf{B}}$ of \mathbf{k} on \mathbf{B} is constant during the motion. The motion in \mathbf{k} space is on a plane normal to the direction of \mathbf{B} , and the orbit is defined by the intersection of this plane with a surface of constant energy.

Physical Derivation of $\hbar\dot{\mathbf{k}} = \mathbf{F}$

We consider the Bloch eigenfunction $\psi_{\mathbf{k}}$ belonging to the energy eigenvalue $\epsilon_{\mathbf{k}}$ and wavevector \mathbf{k} :

$$\psi_{\mathbf{k}} = \sum_{\mathbf{G}} C(\mathbf{k} + \mathbf{G}) \exp[i(\mathbf{k} + \mathbf{G}) \cdot \mathbf{r}] . \quad (8)$$

The expectation value of the momentum of an electron in the Bloch state \mathbf{k} is

$$\mathbf{P}_{\text{el}} = (\mathbf{k} | -i\hbar\nabla | \mathbf{k}) = \sum_{\mathbf{G}} \hbar(\mathbf{k} + \mathbf{G}) |C(\mathbf{k} + \mathbf{G})|^2 = \hbar(\mathbf{k} + \sum_{\mathbf{G}} \mathbf{G} |C(\mathbf{k} + \mathbf{G})|^2) , \quad (9)$$

using $\sum |C(\mathbf{k} + \mathbf{G})|^2 = 1$.

We examine the transfer of momentum between the electron and the lattice when the state \mathbf{k} of the electron is changed to $\mathbf{k} + \Delta\mathbf{k}$ by the application of an external force. We imagine an insulating crystal electrostatically neutral except for a single electron in the state \mathbf{k} of an otherwise empty band.

We suppose that a weak external force is applied for a time interval such that the total impulse given to the entire crystal system is $\mathbf{J} = \int \mathbf{F} dt$. If the conduction electron were free ($m^* = m$), the total momentum imparted to the crystal system by the impulse would appear in the change of momentum of the conduction electron:

$$\mathbf{J} = \Delta\mathbf{p}_{\text{tot}} = \Delta\mathbf{p}_{\text{el}} = \hbar\Delta\mathbf{k} . \quad (10)$$

The neutral crystal suffers no net interaction with the electric field, either directly or indirectly through the free electron.

If the conduction electron interacts with the periodic potential of the crystal lattice, we must have

$$\mathbf{J} = \Delta\mathbf{p}_{\text{tot}} = \Delta\mathbf{p}_{\text{lat}} + \Delta\mathbf{p}_{\text{el}} . \quad (11)$$

From the result (9) for \mathbf{p}_{el} we have

$$\Delta\mathbf{p}_{\text{el}} = \hbar\Delta\mathbf{k} + \sum_{\mathbf{G}} \hbar\mathbf{G} [(\nabla_{\mathbf{k}} |C(\mathbf{k} + \mathbf{G})|^2) \cdot \Delta\mathbf{k}] . \quad (12)$$

The change $\Delta\mathbf{p}_{\text{lat}}$ in the lattice momentum resulting from the change of state of the electron may be derived by an elementary physical consideration. An electron reflected by the lattice transfers momentum to the lattice. If an incident electron with plane wave component of momentum $\hbar\mathbf{k}$ is reflected with momentum $\hbar(\mathbf{k} + \mathbf{G})$, the lattice acquires the momentum $-\hbar\mathbf{G}$, as required by momentum conservation. The momentum transfer to the lattice when the state $\psi_{\mathbf{k}}$ goes over to $\psi_{\mathbf{k}+\Delta\mathbf{k}}$ is

$$\Delta\mathbf{p}_{\text{lat}} = -\hbar \sum_{\mathbf{G}} \mathbf{G} [(\nabla_{\mathbf{k}} |C(\mathbf{k} + \mathbf{G})|^2) \cdot \Delta\mathbf{k}] , \quad (13)$$

because the portion

$$\nabla_{\mathbf{k}}|C(\mathbf{k} + \mathbf{G})|^2 \cdot \Delta\mathbf{k} \quad (14)$$

of each individual component of the initial state is reflected during the state change $\Delta\mathbf{k}$.

The total momentum change is therefore

$$\Delta\mathbf{p}_{\text{el}} + \Delta\mathbf{p}_{\text{lat}} = \mathbf{J} = \hbar\Delta\mathbf{k} , \quad (15)$$

exactly as for free electrons, Eq. (10). Thus from the definition of \mathbf{J} , we have

$$\hbar d\mathbf{k}/dt = \mathbf{F} , \quad (16)$$

derived in (5) by a different method. A rigorous derivation of (16) by an entirely different method is given in Appendix E.

Holes

The properties of vacant orbitals in an otherwise filled band are important in semiconductor physics and in solid state electronics. Vacant orbitals in a band are commonly called holes, and without holes there would be no transistors. A hole acts in applied electric and magnetic fields as if it has a positive charge $+e$. The reason is given in five steps in the boxes that follow.

$$1. \quad \mathbf{k}_h = -\mathbf{k}_e . \quad (17)$$

The total wavevector of the electrons in a filled band is zero: $\Sigma\mathbf{k} = 0$, where the sum is over all states in a Brillouin zone. This result follows from the geometrical symmetry of the Brillouin zone: every fundamental lattice type has symmetry under the inversion operation $\mathbf{r} \rightarrow -\mathbf{r}$ about any lattice point; it follows that the Brillouin zone of the lattice also has inversion symmetry. If the band is filled all pairs of orbitals \mathbf{k} and $-\mathbf{k}$ are filled, and the total wavevector is zero.

If an electron is missing from an orbital of wavevector \mathbf{k}_e , the total wavevector of the system is $-\mathbf{k}_e$ and is attributed to the hole. This result is surprising: the electron is missing from \mathbf{k}_e and the position of the hole is usually indicated graphically as situated at \mathbf{k}_e , as in Fig. 7. But the true wavevector \mathbf{k}_h of the hole is $-\mathbf{k}_e$, which is the wavevector of the point G if the hole is at E . The wavevector $-\mathbf{k}_e$ enters into selection rules for photon absorption.

The hole is an alternate description of a band with one missing electron, and we either say that the hole has wavevector $-\mathbf{k}_e$ or that the band with one missing electron has total wavevector $-\mathbf{k}_e$.

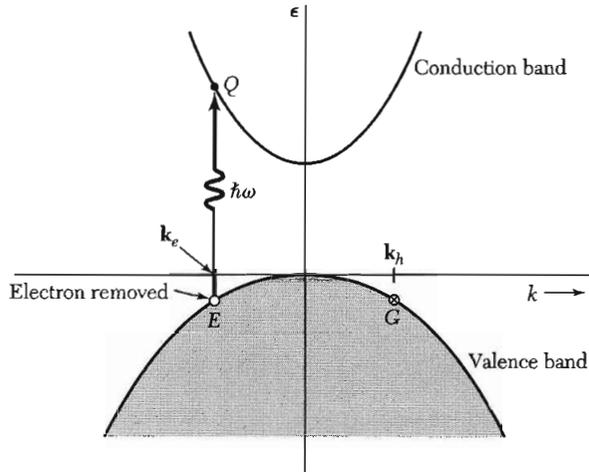


Figure 7 Absorption of a photon of energy $\hbar\omega$ and negligible wavevector takes an electron from E in the filled valence band to Q in the conduction band. If \mathbf{k}_e was the wavevector of the electron at E , it becomes the wavevector of the electron at Q . The total wavevector of the valence band after the absorption is $-\mathbf{k}_e$, and this is the wavevector we must ascribe to the hole if we describe the valence band as occupied by one hole. Thus $\mathbf{k}_h = -\mathbf{k}_e$; the wavevector of the hole is the same as the wavevector of the electron which remains at G . For the entire system the total wavevector after the absorption of the photon is $\mathbf{k}_e + \mathbf{k}_h = 0$, so that the total wavevector is unchanged by the absorption of the photon and the creation of a free electron and free hole.

$$2. \quad \epsilon_h(\mathbf{k}_h) = -\epsilon_e(\mathbf{k}_e) . \quad (18)$$

Here the zero of energy of the valence band is at the top of the band. The lower in the band the missing electron lies, the higher the energy of the system. The energy of the hole is opposite in sign to the energy of the missing electron, because it takes more work to remove an electron from a low orbital than from a high orbital. Thus if the band is symmetric,¹ $\epsilon_e(\mathbf{k}_e) = \epsilon_e(-\mathbf{k}_e) = -\epsilon_h(-\mathbf{k}_e) = -\epsilon_h(\mathbf{k}_h)$. We construct in Fig. 8 a band scheme to represent the properties of a hole. This hole band is a helpful representation because it appears right side up.

$$3. \quad \mathbf{v}_h = \mathbf{v}_e . \quad (19)$$

The velocity of the hole is equal to the velocity of the missing electron. From Fig. 8 we see that $\nabla\epsilon_h(\mathbf{k}_h) = \nabla\epsilon_e(\mathbf{k}_e)$, so that $\mathbf{v}_h(\mathbf{k}_h) = \mathbf{v}_e(\mathbf{k}_e)$.

¹Bands are always symmetric under the inversion $\mathbf{k} \rightarrow -\mathbf{k}$ if the spin-orbit interaction is neglected. Even with spin-orbit interaction, bands are always symmetric if the crystal structure permits the inversion operation. Without a center of symmetry, but with spin-orbit interaction, the bands are symmetric if we compare subbands for which the spin direction is reversed: $\epsilon(\mathbf{k}, \uparrow) = \epsilon(-\mathbf{k}, \downarrow)$. See QTS, Chapter 9.

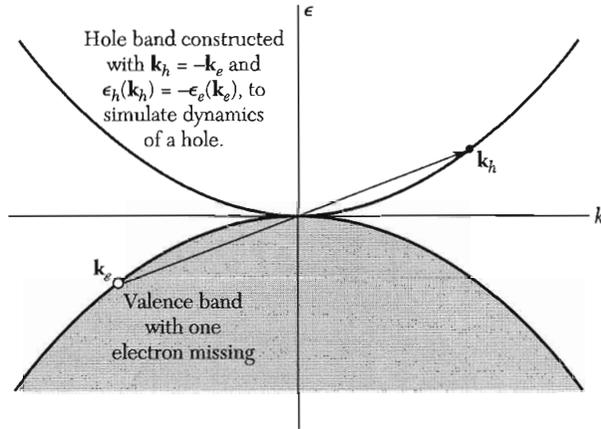


Figure 8 The upper half of the figure shows the hole band that simulates the dynamics of a hole, constructed by inversion of the valence band in the origin. The wavevector and energy of the hole are equal, but opposite in sign, to the wavevector and energy of the empty electron orbital in the valence band. We do not show the disposition of the electron removed from the valence band at \mathbf{k}_e .

$$4. \quad m_h = -m_e . \quad (20)$$

We show below that the effective mass is inversely proportional to the curvature $d^2\epsilon/dk^2$, and for the hole band this has the opposite sign to that for an electron in the valence band. Near the top of the valence band m_e is negative, so that m_h is positive.

$$5. \quad \hbar \frac{d\mathbf{k}_h}{dt} = e(\mathbf{E} + \frac{1}{c} \mathbf{v}_h \times \mathbf{B}) . \quad (21)$$

This comes from the equation of motion

$$(CGS) \quad \hbar \frac{d\mathbf{k}_e}{dt} = -e(\mathbf{E} + \frac{1}{c} \mathbf{v}_e \times \mathbf{B}) \quad (22)$$

that applies to the missing electron when we substitute $-\mathbf{k}_h$ for \mathbf{k}_e and \mathbf{v}_h for \mathbf{v}_e . **The equation of motion for a hole is that of a particle of positive charge e .** The positive charge is consistent with the electric current carried by the valence band of Fig. 9: the current is carried by the unpaired electron in the orbital G :

$$\mathbf{j} = (-e)\mathbf{v}(G) = (-e)[- \mathbf{v}(E)] = e\mathbf{v}(E) , \quad (23)$$

which is just the current of a positive charge moving with the velocity ascribed to the missing electron at E . The current is shown in Fig. 10.

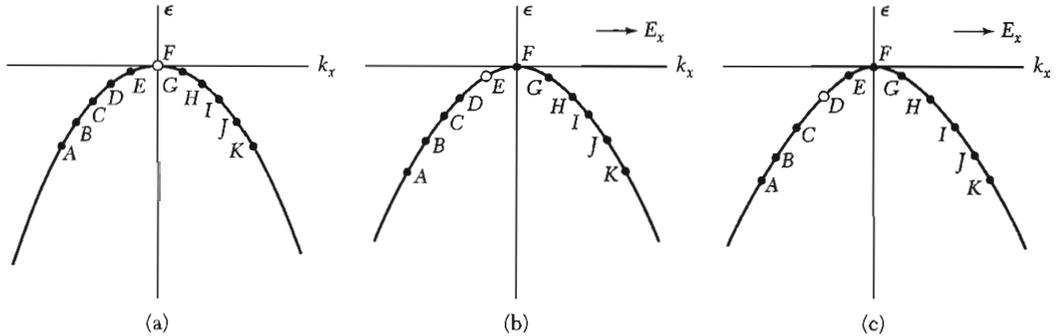


Figure 9 (a) At $t = 0$ all states are filled except F at the top of the band; the velocity v_x is zero at F because $d\epsilon/dk_x = 0$. (b) An electric field E_x is applied in the $+x$ direction. The force on the electrons is in the $-k_x$ direction and all electrons make transitions together in the $-k_x$ direction, moving the hole to the state E . (c) After a further interval the electrons move farther along in k space and the hole is now at D .

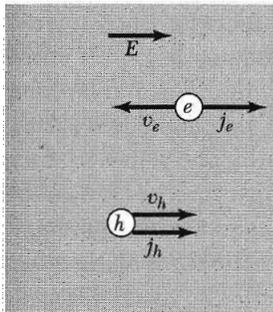


Figure 10 Motion of electrons in the conduction band and holes in the valence band in the electric field E . The hole and electron drift velocities are in opposite directions, but their electric currents are in the same direction, the direction of the electric field.

Effective Mass

When we look at the energy-wavevector relation $\epsilon = (\hbar^2/2m)k^2$ for free electrons, we see that the coefficient of k^2 determines the curvature of ϵ versus k . Turned about, we can say that $1/m$, the reciprocal mass, determines the curvature. For electrons in a band there can be regions of unusually high curvature near the band gap at the zone boundary, as we see from the solutions in Chapter 7 of the wave equation near the zone boundary. If the energy gap is small in comparison with the free electron energy λ at the boundary, the curvature is enhanced by the factor λ/E_g .

In semiconductors the band width, which is like the free electron energy, is of the order of 20 eV, while the band gap is of the order of 0.2 to 2 eV. Thus the reciprocal mass is enhanced by a factor 10 to 100, and the effective mass is reduced to 0.1–0.01 of the free electron mass. These values apply near the band gap; as we go away from the gap the curvatures and the masses are likely to approach those of free electrons.

To summarize the solutions of Chapter 7 for U positive, an electron near the lower edge of the second band has an energy that may be written as

$$\epsilon(K) = \epsilon_c + (\hbar^2/2m_e)K^2; \quad m_e/m = 1/[(2\lambda/U) - 1]. \quad (24)$$

Here K is the wavevector measured from the zone boundary, and m_e denotes the effective mass of the electron near the edge of the second band. An electron near the top of the first band has the energy

$$\epsilon(K) = \epsilon_0 - (\hbar^2/2m_h)K^2 ; \quad m_h/m = 1/[(2\lambda/U) + 1] . \quad (25)$$

The curvature and hence the mass will be negative near the top of the first band, but we have introduced a minus sign into (25) in order that the symbol m_h for the hole mass will have a positive value—see (20) above.

The crystal does not weigh any less if the effective mass of a carrier is less than the free electron mass, nor is Newton's second law violated for the crystal *taken as a whole*, ions plus carriers. The important point is that an electron in a periodic potential is accelerated relative to the lattice in an applied electric or magnetic field as if the mass of the electron were equal to an effective mass which we now define.

We differentiate the result (1) for the group velocity to obtain

$$\frac{dv_g}{dt} = \hbar^{-1} \frac{d^2\epsilon}{dk dt} = \hbar^{-1} \left(\frac{d^2\epsilon}{dk^2} \frac{dk}{dt} \right) . \quad (26)$$

We know from (5) that $dk/dt = F/\hbar$, whence

$$\frac{dv_g}{dt} = \left(\frac{1}{\hbar^2} \frac{d^2\epsilon}{dk^2} \right) F ; \quad \text{or} \quad F = \frac{\hbar^2}{d^2\epsilon/dk^2} \frac{dv_g}{dt} . \quad (27)$$

If we identify $\hbar^2/(d^2\epsilon/dk^2)$ as a mass, then (27) assumes the form of Newton's second law. We define the **effective mass** m^* by

$$\boxed{\frac{1}{m^*} = \frac{1}{\hbar^2} \frac{d^2\epsilon}{dk^2}} . \quad (28)$$

It is easy to generalize this to take account of an anisotropic electron energy surface, as for electrons in Si or Ge. We introduce the components of the reciprocal effective mass tensor

$$\left(\frac{1}{m^*} \right)_{\mu\nu} = \frac{1}{\hbar^2} \frac{d^2\epsilon_k}{dk_\mu dk_\nu} ; \quad \frac{dv_\mu}{dt} = \left(\frac{1}{m^*} \right)_{\mu\nu} F_\nu , \quad (29)$$

where μ, ν are Cartesian coordinates.

Physical Interpretation of the Effective Mass

How can an electron of mass m when put into a crystal respond to applied fields as if the mass were m^* ? It is helpful to think of the process of Bragg reflection of electron waves in a lattice. Consider the weak interaction approximation treated in Chapter 7. Near the bottom of the lower band the orbital is represented quite adequately by a plane wave $\exp(ikx)$ with momentum $\hbar k$; the wave component $\exp[i(k-G)x]$ with momentum $\hbar(k-G)$ is small and

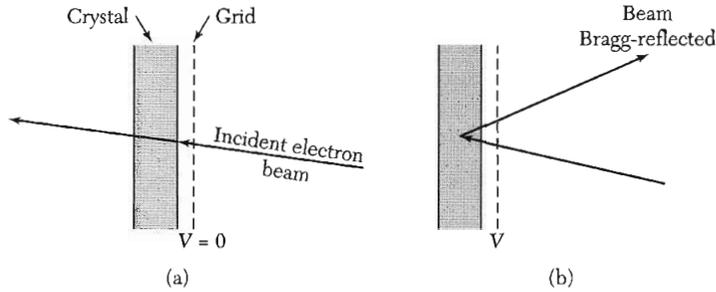


Figure 11 Explanation of negative effective masses which occur near, but below, a Brillouin zone boundary. In (a) the energy of the electron beam incident on a thin crystal is slightly too low to satisfy the condition for Bragg reflection and the beam is transmitted through the crystal. The application of a small voltage across the grid may, as in (b), cause the Bragg condition to be satisfied, and the electron beam will then be reflected from the appropriate set of crystal planes.

increases only slowly as k is increased, and in this region $m^* \approx m$. An increase in the reflected component $\exp[i(k - G)x]$ as k is increased represents momentum transfer to the electron from the lattice.

Near the boundary the reflected component is quite large; at the boundary it becomes equal in amplitude to the forward component, at which point the eigenfunctions are standing waves, rather than running waves. Here the momentum component $\hbar(-\frac{1}{2}G)$ cancels the momentum component $\hbar(\frac{1}{2}G)$.

A single electron in an energy band may have positive or negative effective mass: the states of positive effective mass occur near the bottom of a band because positive effective mass means that the band has upward curvature ($d^2\epsilon/dk^2$ is positive). States of negative effective mass occur near the top of the band. A negative effective mass means that on going from state k to state $k + \Delta k$, the momentum transfer to the lattice from the electron is larger than the momentum transfer from the applied force to the electron. Although k is increased by Δk by the applied electric field, the approach to Bragg reflection can give an overall decrease in the forward momentum of the electron; when this happens the effective mass is negative (Fig. 11).

As we proceed in the second band away from the boundary, the amplitude of $\exp[i(k - G)x]$ decreases rapidly and m^* assumes a small positive value. Here the increase in electron velocity resulting from a given external impulse is larger than that which a free electron would experience. The lattice makes up the difference through the reduced recoil it experiences when the amplitude of $\exp[i(k - G)x]$ is diminished.

If the energy in a band depends only slightly on k , then the effective mass will be very large. That is, $m^*/m \gg 1$ when $d^2\epsilon/dk^2$ is very small. The tight-binding approximation discussed in Chapter 9 gives quick insight into the formation of narrow bands. If the wavefunctions centered on neighboring atoms overlap very little, then the overlap integral is small; the width of the band

narrow, and the effective mass large. The overlap of wavefunctions centered on neighboring atoms is small for the inner or core electrons. The $4f$ electrons of the rare earth metals, for example, overlap very little.

Effective Masses in Semiconductors

In many semiconductors it has been possible to determine by cyclotron resonance the effective masses of carriers in the conduction and valence bands near the band edges. The determination of the energy surface is equivalent to a determination of the effective mass tensor (29). Cyclotron resonance in a semiconductor is carried out with centimeter wave or millimeter wave radiation at low carrier concentration.

The current carriers are accelerated in helical orbits about the axis of a static magnetic field. The angular rotation frequency ω_c is

$$\text{(CGS)} \quad \omega_c = \frac{eB}{m^*c}, \quad \text{(SI)} \quad \omega_c = \frac{eB}{m^*} \quad (30)$$

where m^* is the appropriate cyclotron effective mass. Resonant absorption of energy from an rf electric field perpendicular to the static magnetic field (Fig. 12) occurs when the rf frequency is equal to the cyclotron frequency. Holes and electrons rotate in opposite senses in a magnetic field.

We consider the experiment for $m^*/m = 0.1$. At $f_c = 24$ GHz, or $\omega_c = 1.5 \times 10^{11} \text{ s}^{-1}$, we have $B = 860$ G at resonance. The line width is determined by the collision relaxation time τ , and to obtain a distinctive resonance it is necessary that $\omega_c \tau \geq 1$. The mean free path must be long enough to permit the average carrier to get one radian around a circle between collisions. The requirements are met with the use of higher frequency radiation and higher magnetic fields, with high purity crystals in liquid helium.

In direct-gap semiconductors with band edges at the center of the Brillouin zone, the bands have the structure shown in Fig. 13. The conduction band edge is spherical with the effective mass m_0 :

$$\epsilon_c = E_g + \hbar^2 k^2 / 2m_e, \quad (31)$$

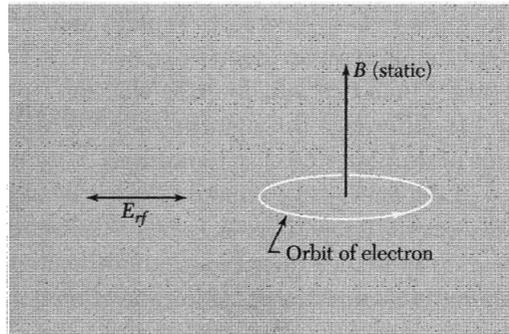


Figure 12 Arrangement of fields in a cyclotron resonance experiment in a semiconductor. The sense of the circulation is opposite for electrons and holes.

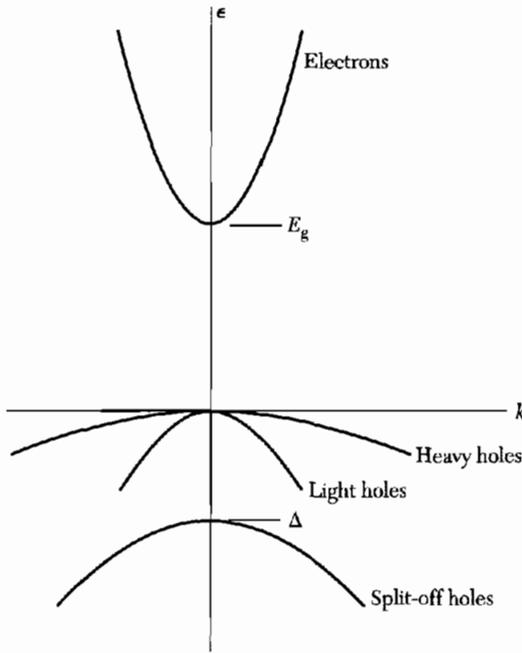


Figure 13 Simplified view of the band edge structure of a direct-gap semiconductor.

Table 2 Effective masses of electrons and holes in direct-gap semiconductors

Crystal	Electron m_e/m	Heavy hole m_{hh}/m	Light hole m_{lh}/m	Split-off hole m_{soh}/m	Spin-orbit Δ , eV
InSb	0.015	0.39	0.021	(0.11)	0.82
InAs	0.026	0.41	0.025	0.08	0.43
InP	0.073	0.4	(0.078)	(0.15)	0.11
GaSb	0.047	0.3	0.06	(0.14)	0.80
GaAs	0.066	0.5	0.082	0.17	0.34
Cu ₂ O	0.99	—	0.58	0.69	0.13

referred to the valence band edge. The valence bands are characteristically threefold near the edge, with the heavy hole hh and light hole lh bands degenerate at the center, and a band soh split off by the spin-orbit splitting Δ :

$$\begin{aligned} \epsilon_v(hh) &\cong -\hbar^2 k^2 / 2m_{hh} ; & \epsilon_v(lh) &\cong -\hbar^2 k^2 / 2m_{lh} ; \\ \epsilon_v(soh) &\cong -\Delta - \hbar^2 k^2 / 2m_{soh} . \end{aligned} \quad (32)$$

Values of the mass parameters are given in Table 2. The forms (32) are only approximate, because even close to $k = 0$ the heavy and light hole bands are not spherical—see the discussion below for Ge and Si.

The perturbation theory of band edges (Problem 9.8) suggests that the electron effective mass should be proportional to the band gap, approximately,

for a direct gap crystal. We use Tables 1 and 2 to find the fairly constant values $m_e/(mE_g) = 0.063, 0.060, \text{ and } 0.051$ in $(\text{eV})^{-1}$ for the series InSb, InAs, and InP, in agreement with this suggestion.

Silicon and Germanium

The conduction and valence bands of germanium are shown in Fig. 14, based on a combination of theoretical and experimental results. The valence band edge in both Si and Ge is at $\mathbf{k} = 0$ and is derived from $p_{3/2}$ and $p_{1/2}$ states of the free atoms, as is clear from the tight-binding approximation (Chapter 9) to the wavefunctions.

The $p_{3/2}$ level is fourfold degenerate as in the atom; the four states correspond to m_j values $\pm \frac{3}{2}$ and $\pm \frac{1}{2}$. The $p_{1/2}$ level is doubly degenerate, with $m_j = \pm \frac{1}{2}$. The $p_{3/2}$ states are higher in energy than the $p_{1/2}$ states; the energy difference Δ is a measure of the spin-orbit interaction.

The valence band edges are not simple. Holes near the band edge are characterized by two effective masses, light and heavy. These arise from the two bands formed from the $p_{3/2}$ level of the atom. There is also a band formed from the $p_{1/2}$ level, split off from the $p_{3/2}$ level by the spin-orbit interaction. The energy surfaces are not spherical, but warped (QTS, p. 271):

$$\epsilon(\mathbf{k}) = Ak^2 \pm [B^2k^4 + C^2(k_x^2k_y^2 + k_y^2k_z^2 + k_z^2k_x^2)]^{1/2} \quad (33)$$

The choice of sign distinguishes the two masses. The split-off band has $\epsilon(k) = -\Delta + Ak^2$. The experiments give, in units $\hbar^2/2m$,

$$\begin{array}{llll} \text{Si:} & A = -4.29 & ; & |B| = 0.68 & ; & |C| = 4.87 & ; & \Delta = 0.044 \text{ eV} \\ \text{Ge:} & A = -13.38 & ; & |B| = 8.48 & ; & |C| = 13.15 & ; & \Delta = 0.29 \text{ eV} \end{array}$$

Roughly, the light and heavy holes in germanium have masses $0.043 m$ and $0.34 m$; in silicon $0.16 m$ and $0.52 m$; in diamond $0.7 m$ and $2.12 m$.

The conduction band edges in Ge are at the equivalent points L of the Brillouin zone, Fig. 15a. Each band edge has a spheroidal energy surface oriented along a $\langle 111 \rangle$ crystal axis, with a longitudinal mass $m_l = 1.59 m$ and a transverse mass $m_t = 0.082 m$. For a static magnetic field at an angle θ with the longitudinal axis of a spheroid, the effective cyclotron mass m_c is

$$\frac{1}{m_c} = \frac{\cos^2\theta}{m_t} + \frac{\sin^2\theta}{m_l m_t} \quad (34)$$

Results for Ge are shown in Fig. 16.

In silicon the conduction band edges are spheroids oriented along the equivalent $\langle 100 \rangle$ directions in the Brillouin zone, with mass parameters $m_l = 0.92 m$ and $m_t = 0.19 m$, as in Fig. 17a. The band edges lie along the lines labeled Δ in the zone of Fig. 15a, a little way in from the boundary points X .

In GaAs we have $A = -6.98$, $B = -4.5$, $|C| = 6.2$, $\Delta = 0.341 \text{ eV}$. The band structure is shown in Fig. 17b. It has a direct band gap with an isotropic conduction electron mass of $0.067 m$.

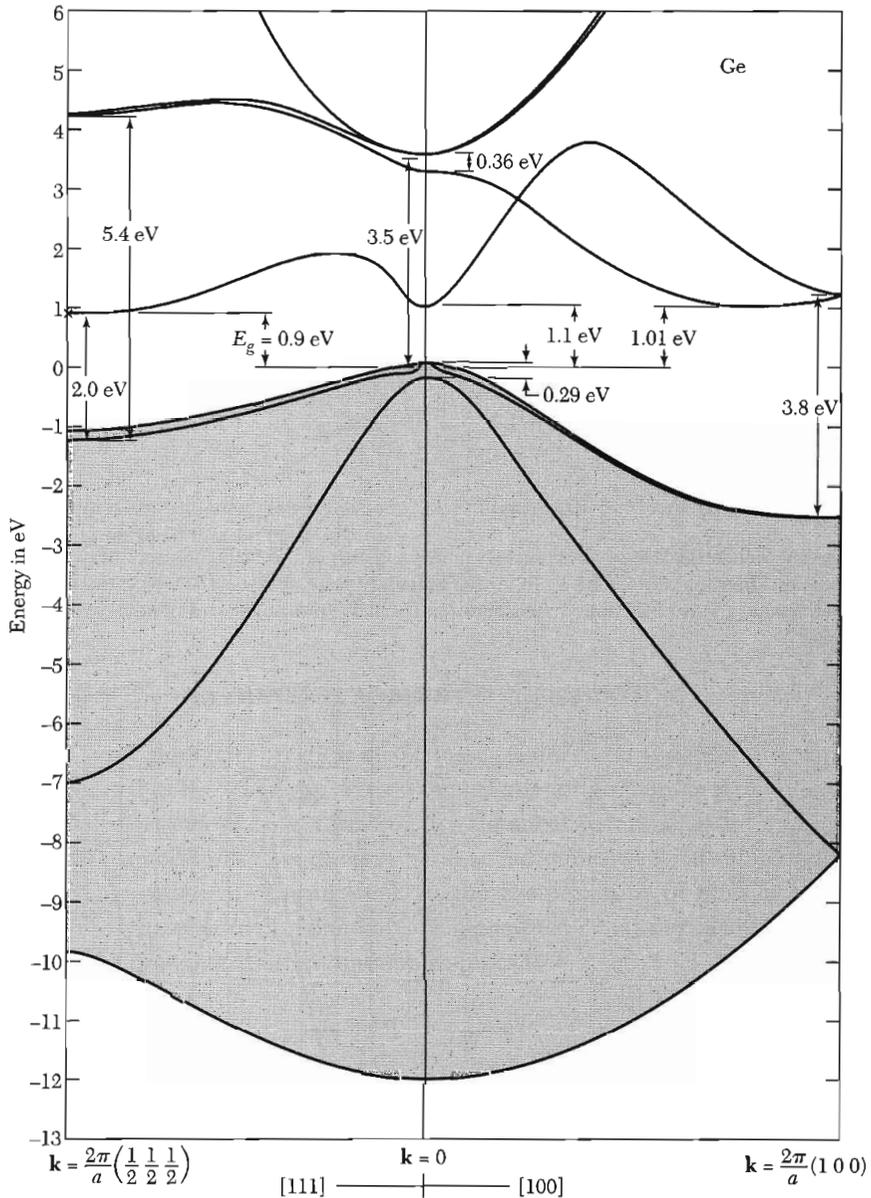


Figure 14 Calculated band structure of germanium, after C. Y. Fong. The general features are in good agreement with experiment. The four valence bands are shown in gray. The fine structure of the valence band edge is caused by spin-orbit splitting. The energy gap is indirect; the conduction band edge is at the point $(2\pi/a)(\frac{1}{2} \frac{1}{2} \frac{1}{2})$. The constant energy surfaces around this point are ellipsoidal.

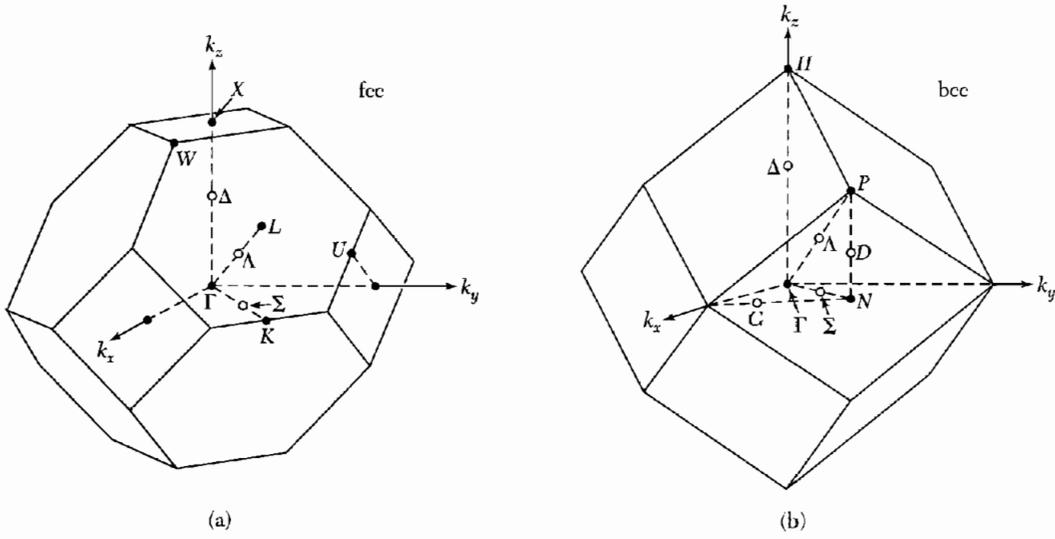


Figure 15 Standard labels of the symmetry points and axes of the Brillouin zones of the fcc and bcc lattices. The zone centers are Γ . In (a) the boundary point at $(2\pi/a)(100)$ is X ; the boundary point at $(2\pi/a)(\frac{1}{2} \frac{1}{2} \frac{1}{2})$ is L ; the line Δ runs between Γ and X . In (b) the corresponding symbols are H , P , and Δ .

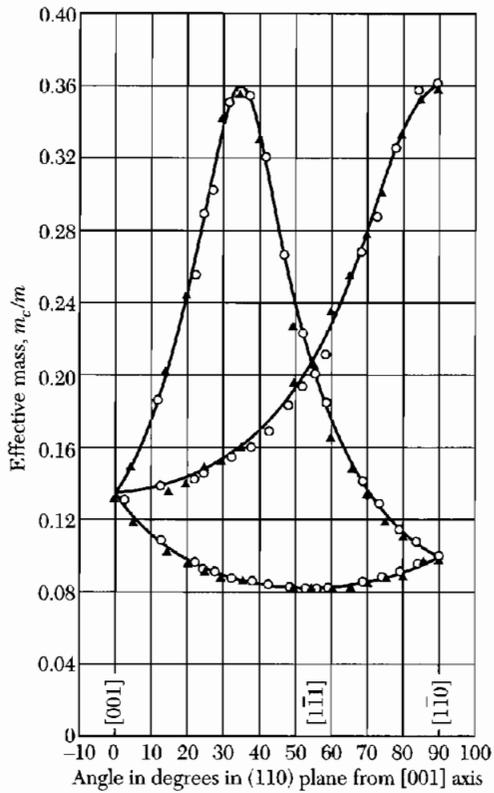
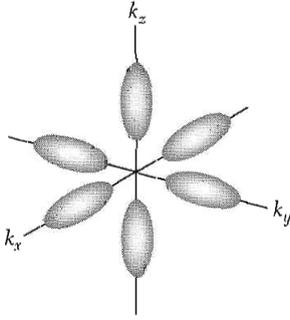
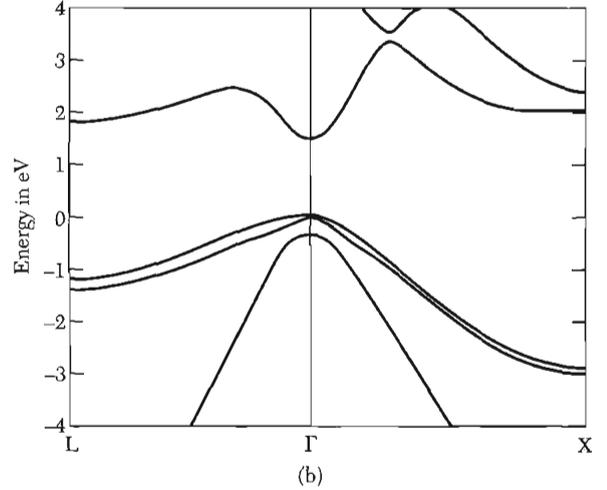


Figure 16 Effective cyclotron mass of electrons in germanium at 4 K for magnetic field directions in a (110) plane. There are four independent mass spheroids in Ge, one along each $[111]$ axis, but viewed in the (110) plane two spheroids always appear equivalent. (After Dresselhaus, Kip, and Kittel.)



(a)

Figure 17a Constant energy ellipsoids for electrons in silicon, drawn for $m_l/m_e = 5$.



(b)

Figure 17b Band structure of GaAs, after S. G. Louie.

INTRINSIC CARRIER CONCENTRATION

We want the concentration of intrinsic carriers as a function of temperature, in terms of the band gap. We do the calculation for simple parabolic band edges. We first calculate in terms of the chemical potential μ the number of electrons excited to the conduction band at temperature T . In semiconductor physics μ is called the Fermi level. At the temperatures of interest we may suppose for the conduction band of a semiconductor that $\epsilon - \mu \gg k_B T$, so that the Fermi-Dirac distribution function reduces to

$$f_e \approx \exp[(\mu - \epsilon)/k_B T] . \quad (35)$$

This is the probability that a conduction electron orbital is occupied, in an approximation valid when $f_e \ll 1$.

The energy of an electron in the conduction band is

$$\epsilon_k = E_c + \hbar^2 k^2 / 2m_e , \quad (36)$$

where E_c is the energy at the conduction band edge, as in Fig. 18. Here m_e is the effective mass of an electron. Thus from (6.20) the density of states at ϵ is

$$D_e(\epsilon) = \frac{1}{2\pi^2} \left(\frac{2m_e}{\hbar^2} \right)^{3/2} (\epsilon - E_c)^{1/2} . \quad (37)$$

The concentration of electrons in the conduction band is

$$n = \int_{E_c}^{\infty} D_c(\epsilon) f_e(\epsilon) d\epsilon = \frac{1}{2\pi^2} \left(\frac{2m_e}{\hbar^2} \right)^{3/2} \exp(\mu/k_B T) \times \int_{E_c}^{\infty} (\epsilon - E_c)^{1/2} \exp(-\epsilon/k_B T) d\epsilon, \quad (38)$$

which integrates to give

$$n = 2 \left(\frac{m_e k_B T}{2\pi\hbar^2} \right)^{3/2} \exp[(\mu - E_c)/k_B T]. \quad (39)$$

The problem is solved for n when μ is known. It is useful to calculate the equilibrium concentration of holes p . The distribution function f_h for holes is related to the electron distribution function f_e by $f_h = 1 - f_e$, because a hole is the absence of an electron. We have

$$f_h = 1 - \frac{1}{\exp[(\epsilon - \mu)/k_B T] + 1} = \frac{1}{\exp[(\mu - \epsilon)/k_B T] + 1} \cong \exp[(\epsilon - \mu)/k_B T], \quad (40)$$

provided $(\mu - \epsilon) \gg k_B T$.

If the holes near the top of the valence band behave as particles with effective mass m_h , the density of hole states is given by

$$D_h(\epsilon) = \frac{1}{2\pi^2} \left(\frac{2m_h}{\hbar^2} \right)^{3/2} (E_v - \epsilon)^{1/2}, \quad (41)$$

where E_v is the energy at the valence band edge. Proceeding as in (38) we obtain

$$p = \int_{-\infty}^{E_c} D_h(\epsilon) f_h(\epsilon) d\epsilon = 2 \left(\frac{m_h k_B T}{2\pi\hbar^2} \right)^{3/2} \exp[(E_c - \mu)/k_B T] \quad (42)$$

for the concentration p of holes in the valence band.

We multiply together the expressions for n and p to obtain the equilibrium relation, with the energy gap $E_g = E_c - E_v$ as in Fig. 18,

$$np = 4 \left(\frac{k_B T}{2\pi\hbar^2} \right)^3 (m_e m_h)^{3/2} \exp(-E_g/k_B T). \quad (43)$$

This useful result does not involve the Fermi level μ . At 300 K the value of np is $2.10 \times 10^{19} \text{ cm}^{-6}$, $2.89 \times 10^{26} \text{ cm}^{-6}$, and $6.55 \times 10^{12} \text{ cm}^{-6}$, for the actual band structures of Si, Ge, and GaAs, respectively.

We have nowhere assumed in the derivation that the material is intrinsic: the result holds for impurity ionization as well. The only assumption made is that the distance of the Fermi level from the edge of both bands is large in comparison with $k_B T$.

A simple kinetic argument shows why the product np is constant at a given temperature. Suppose that the equilibrium population of electrons and holes

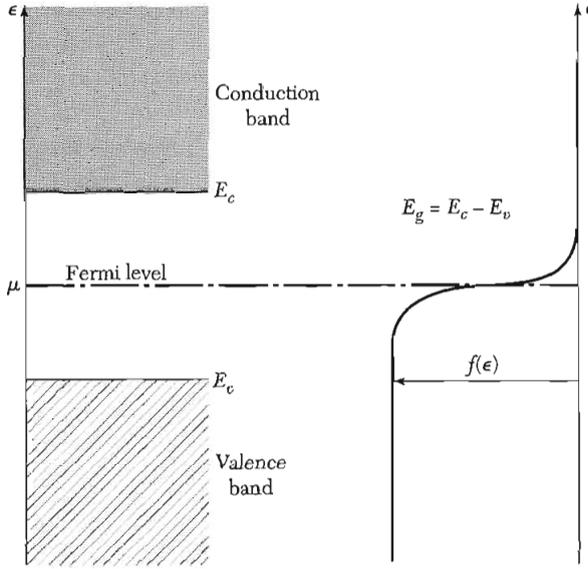


Figure 18 Energy scale for statistical calculations. The Fermi distribution function is shown on the same scale, for a temperature $k_B T \ll E_g$. The Fermi level μ is taken to lie well within the band gap, as for an intrinsic semiconductor. If $\epsilon = \mu$, then $f = \frac{1}{2}$.

is maintained by black-body photon radiation at temperature T . The photons generate electron-hole pairs at a rate $A(T)$, while $B(T)np$ is the rate of the recombination reaction $e + h = \text{photon}$. Then

$$dn/dt = A(T) - B(T)np = dp/dt \quad (44)$$

In equilibrium $dn/dt = 0$, $dp/dt = 0$, whence $np = A(T)/B(T)$.

Because the product of the electron and hole concentrations is a constant independent of impurity concentration at a given temperature, the introduction of a small proportion of a suitable impurity to increase n , say, must decrease p . This result is important in practice—we can reduce the total carrier concentration $n + p$ in an impure crystal, sometimes enormously, by the controlled introduction of suitable impurities. Such a reduction is called **compensation**.

In an intrinsic semiconductor the number of electrons is equal to the number of holes, because the thermal excitation of an electron leaves behind a hole in the valence band. Thus from (43) we have, letting the subscript i denote intrinsic and $E_g = E_c - E_v$,

$$n_i = p_i = 2 \left(\frac{k_B T}{2\pi\hbar^2} \right)^{3/2} (m_e m_h)^{3/4} \exp(-E_g/2k_B T) \quad (45)$$

The intrinsic carrier concentration depends exponentially on $E_g/2k_B T$, where E_g is the energy gap. We set (39) equal to (42) to obtain, for the Fermi level as measured from the top of the valence band,

$$\exp(2\mu/k_B T) = (m_h/m_e)^{3/2} \exp(E_g/k_B T) \quad (46)$$

$$\mu = \frac{1}{2} E_g + \frac{3}{4} k_B T \ln(m_h/m_e) \quad (47)$$

If $m_h = m_e$, then $\mu = \frac{1}{2}E_g$ and the Fermi level is in the middle of the forbidden gap.

Intrinsic Mobility

The mobility is the magnitude of the drift velocity of a charge carrier per unit electric field:

$$\mu = |v|/E . \quad (48)$$

The mobility is defined to be positive for both electrons and holes, although their drift velocities are opposite in a given field. By writing μ_e or μ_h with subscripts for the electron or hole mobility we can avoid any confusion between μ as the chemical potential and as the mobility.

The electrical conductivity is the sum of the electron and hole contributions:

$$\sigma = (ne\mu_e + pe\mu_h) , \quad (49)$$

where n and p are the concentrations of electrons and holes. In Chapter 6 the drift velocity of a charge q was found to be $v = q\tau E/m$, whence

$$\mu_e = e\tau_e/m_e ; \quad \mu_h = e\tau_h/m_h , \quad (50)$$

where τ is the collision time.

The mobilities depend on temperature as a modest power law. The temperature dependence of the conductivity in the intrinsic region will be dominated by the exponential dependence $\exp(-E_g/2k_B T)$ of the carrier concentration, Eq. (45).

Table 3 gives experimental values of the mobility at room temperature. The mobility in SI units is expressed in $\text{m}^2/\text{V}\cdot\text{s}$ and is 10^{-4} of the mobility in practical units. For most substances the values quoted are limited by the scattering of carriers by thermal phonons. The hole mobilities typically are smaller than the electron mobilities because of the occurrence of band degeneracy at the valence band edge at the zone center, thereby making possible interband scattering processes that reduce the mobility considerably.

Table 3 Carrier mobilities at room temperature, in $\text{cm}^2/\text{V}\cdot\text{s}$

Crystal	Electrons	Holes	Crystal	Electrons	Holes
Diamond	1800	1200	GaAs	8000	300
Si	1350	480	GaSb	5000	1000
Ge	3600	1800	PbS	550	600
InSb	800	450	PhSe	1020	930
InAs	30000	450	PbTe	2500	1000
InP	4500	100	AgCl	50	—
AlAs	280	—	KBr (100 K)	100	—
AlSb	900	400	SiC	100	10–20

In some crystals, particularly in ionic crystals, the holes are essentially immobile and get about only by thermally-activated hopping from ion to ion. The principal cause of this "self-trapping" is the lattice distortion associated with the Jahn-Teller effect of degenerate states. The orbital degeneracy necessary for self-trapping is much more frequent for holes than for electrons.

There is a tendency for crystals with small energy gaps at direct band edges to have high values of the electron mobility. Small gaps lead to small effective masses, which favor high mobilities. The highest mobility observed in a bulk semiconductor is 5×10^6 cm²/V-s in PbTe at 4 K, where the gap is 0.19 eV.

IMPURITY CONDUCTIVITY

Certain impurities and imperfections drastically affect the electrical properties of a semiconductor. The addition of boron to silicon in the proportion of 1 boron atom to 10^5 silicon atoms increases the conductivity of pure silicon at room temperature by a factor of 10^3 . In a compound semiconductor a stoichiometric deficiency of one constituent will act as an impurity; such semiconductors are known as **deficit semiconductors**. The deliberate addition of impurities to a semiconductor is called **doping**.

We consider the effect of impurities in silicon and germanium. These elements crystallize in the diamond structure. Each atom forms four covalent bonds, one with each of its nearest neighbors, corresponding to the chemical valence four. If an impurity atom of valence five, such as phosphorus, arsenic, or antimony, is substituted in the lattice in place of a normal atom, there will be one valence electron from the impurity atom left over after the four covalent bonds are established with the nearest neighbors, that is, after the impurity atom has been accommodated in the structure with as little disturbance as possible. Impurity atoms that can give up an electron are called **donors**.

Donor States. The structure in Fig. 19 has a positive charge on the impurity atom (which has lost one electron). Lattice constant studies have verified that the pentavalent impurities enter the lattice by substitution for normal atoms, and not in interstitial positions. The crystal as a whole remains neutral because the electron remains in the crystal.

The extra electron moves in the coulomb potential $e/\epsilon r$ of the impurity ion, where ϵ in a covalent crystal is the static dielectric constant of the medium. The factor $1/\epsilon$ takes account of the reduction in the coulomb force between charges caused by the electronic polarization of the medium. This treatment is valid for orbits large in comparison with the distance between atoms, and for slow motions of the electron such that the orbital frequency is low in comparison with the frequency ω_g corresponding to the energy gap. These conditions are satisfied quite well in Ge and Si by the donor electron of P, As, or Sb.

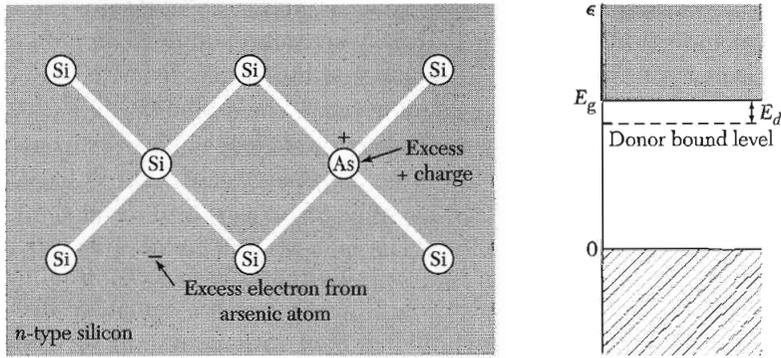


Figure 19 Charges associated with an arsenic impurity atom in silicon. Arsenic has five valence electrons, but silicon has only four valence electrons. Thus four electrons on the arsenic form tetrahedral covalent bonds similar to silicon, and the fifth electron is available for conduction. The arsenic atom is called a **donor** because when ionized it donates an electron to the conduction band.

We estimate the ionization energy of the donor impurity. The Bohr theory of the hydrogen atom may be modified to take into account the dielectric constant of the medium and the effective mass of an electron in the periodic potential of the crystal. The ionization energy of atomic hydrogen is $-e^4m/2\hbar^2$ in CGS and $-e^4m/2(4\pi\epsilon_0\hbar)^2$ in SI.

In the semiconductor with dielectric constant ϵ we replace e^2 by e^2/ϵ and m by the effective mass m_e to obtain

$$\text{(CGS)} \quad E_d = \frac{e^4 m_e}{2\epsilon^2 \hbar^2} = \left(\frac{13.6 m_e}{\epsilon^2 m} \right) eV ; \quad \text{(SI)} \quad E_d = \frac{e^4 m_e}{2(4\pi\epsilon\epsilon_0\hbar)^2} \quad (51)$$

as the donor ionization energy of the semiconductor.

The Bohr radius of the ground state of hydrogen is \hbar^2/me^2 in CGS or $4\pi\epsilon_0\hbar^2/me^2$ in SI. Thus the Bohr radius of the donor is

$$\text{(CGS)} \quad a_d = \frac{\epsilon\hbar^2}{m_e e^2} = \left(\frac{0.53\epsilon}{m_e/m} \right) \text{\AA} ; \quad \text{(SI)} \quad a_d = \frac{4\pi\epsilon\epsilon_0\hbar^2}{m_e e^2} \quad (52)$$

The application of impurity state theory to germanium and silicon is complicated by the anisotropic effective mass of the conduction electrons. But the dielectric constant has the more important effect on the donor energy because it enters as the square, whereas the effective mass enters only as the first power.

To obtain a general impression of the impurity levels we use $m_e \approx 0.1 m$ for electrons in germanium and $m_e \approx 0.2 m$ in silicon. The static dielectric constant is given in Table 4. The ionization energy of the free hydrogen atom is 13.6 eV. For germanium the donor ionization energy E_d on our model is 5 meV, reduced with respect to hydrogen by the factor $m_e/m\epsilon^2 = 4 \times 10^{-4}$. The corresponding result for silicon is 20 meV. Calculations using the correct

Table 4 Static relative dielectric constant of semiconductors

Crystal	ϵ	Crystal	ϵ
Diamond	5.5	GaSb	15.69
Si	11.7	GaAs	13.13
Ge	15.8	AlAs	10.1
InSb	17.88	AlSb	10.3
InAs	14.55	SiC	10.2
InP	12.37	Cu ₂ O	7.1

Table 5 Donor ionization energies E_d of pentavalent impurities in germanium and silicon, in meV

	P	As	Sb
Si	45.	49.	39.
Ge	12.0	12.7	9.6

anisotropic mass tensor predict 9.05 meV for germanium and 29.8 meV for silicon. Observed values of donor ionization energies in Si and Ge are given in Table 5. In GaAs donors have $E_d \approx 6$ meV.

The radius of the first Bohr orbit is increased by $\epsilon m/m_e$ over the value 0.53 \AA for the free hydrogen atom. The corresponding radius is $(160)(0.53) \approx 80 \text{ \AA}$ in germanium and $(60)(0.53) \approx 30 \text{ \AA}$ in silicon. These are large radii, so that donor orbits overlap at relatively low donor concentrations, compared to the number of host atoms. With appreciable orbit overlap, an "impurity band" is formed from the donor states; see the discussion of the metal-insulator transition in Chapter 14.

The semiconductor can conduct in the impurity band by electrons hopping from donor to donor. The process of impurity band conduction sets in at lower donor concentration levels if there are also some acceptor atoms present, so that some of the donors are always ionized. It is easier for a donor electron to hop to an ionized (unoccupied) donor than to an occupied donor atom, in order that two electrons will not have to occupy the same site during charge transport.

Acceptor States. A hole may be bound to a trivalent impurity in germanium or silicon (Fig. 20), just as an electron is bound to a pentavalent impurity. Trivalent impurities such as B, Al, Ga, and In are called **acceptors** because they accept electrons from the valence band in order to complete the covalent bonds with neighbor atoms, leaving holes in the band.

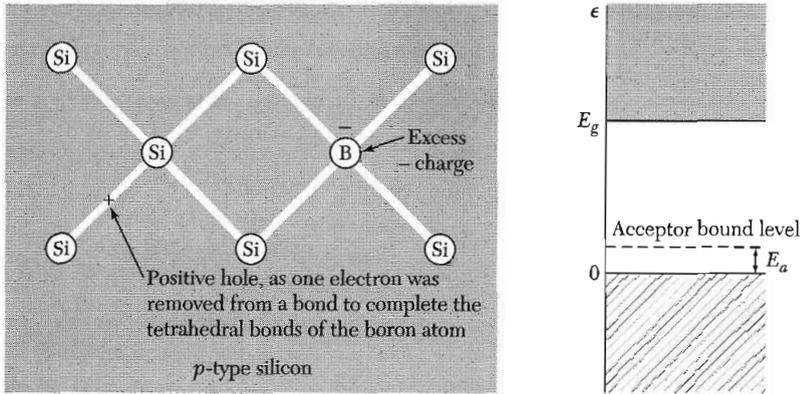


Figure 20 Boron has only three valence electrons; it can complete its tetrahedral bonds only by taking an electron from a Si-Si bond, leaving behind a hole in the silicon valence band. The positive hole is then available for conduction. The boron atom is called an **acceptor** because when ionized it accepts an electron from the valence band. At 0 K the hole is bound.

Table 6 Acceptor ionization energies E_a of trivalent impurities in germanium and silicon, in meV

	B	Al	Ga	In
Si	45.	57.	65.	157.
Ge	10.4	10.2	10.8	11.2

When an acceptor is ionized a hole is freed, which requires an input of energy. On the usual energy band diagram, an electron rises when it gains energy, whereas a hole sinks in gaining energy.

Experimental ionization energies of acceptors in germanium and silicon are given in Table 6. The Bohr model applies qualitatively for holes just as for electrons, but the degeneracy at the top of the valence band complicates the effective mass problem.

The tables show that donor and acceptor ionization energies in Si are comparable with $k_B T$ at room temperature (26 meV), so that the thermal ionization of donors and acceptors is important in the electrical conductivity of silicon at room temperature. If donor atoms are present in considerably greater numbers than acceptors, the thermal ionization of donors will release electrons into the conduction band. The conductivity of the specimen then will be controlled by electrons (negative charges), and the material is said to be *n* type.

If acceptors are dominant, holes will be released into the valence band and the conductivity will be controlled by holes (positive charges): the material is *p* type. The sign of the Hall voltage (6.53) is a rough test for *n* or *p* type.

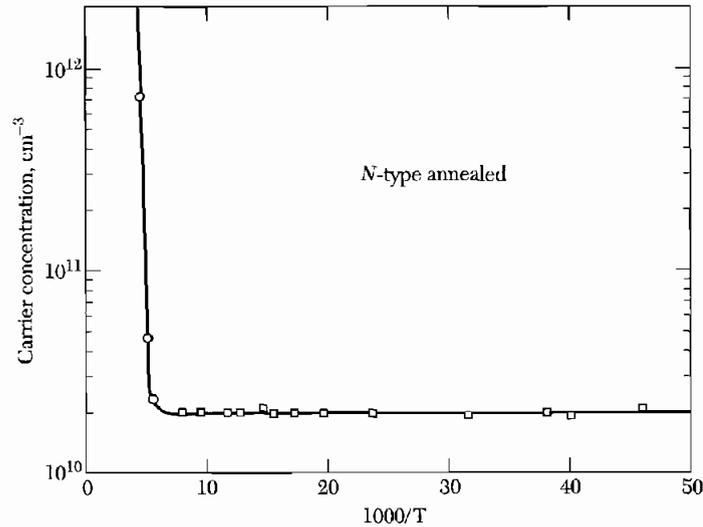


Figure 21 Temperature dependence of the free carrier concentration in ultrapurified Ge, after R. N. Hall. The net concentration of electrically active impurities is $2 \times 10^{10} \text{ cm}^{-3}$, as determined by Hall coefficient measurements. The rapid onset of intrinsic excitation as the temperature is increased is evident at low values of $1/T$. The carrier concentration is closely constant between 20 K and 200 K.

Another handy laboratory test is the sign of the thermoelectric potential, discussed below.

The numbers of holes and electrons are equal in the intrinsic regime. The intrinsic electron concentration n_i at 300 K is $1.7 \times 10^{13} \text{ cm}^{-3}$ in germanium and $4.6 \times 10^9 \text{ cm}^{-3}$ in silicon. The electrical resistivity of intrinsic material is 43 ohm-cm for germanium and 2.6×10^5 ohm-cm for silicon.

Germanium has 4.42×10^{22} atoms per cm^3 . The purification of Ge has been carried further than any other element. The concentration of the common electrically active impurities—the shallow donor and acceptor impurities—has been reduced below 1 impurity atom in 10^{11} Ge atoms (Fig. 21). For example, the concentration of P in Ge can be reduced below $4 \times 10^{10} \text{ cm}^{-3}$. There are impurities (H, O, Si, C) whose concentrations in Ge cannot usually be reduced below 10^{12} – 10^{14} cm^{-3} , but these do not affect electrical measurements and therefore may be hard to detect.

Thermal Ionization of Donors and Acceptors

The calculation of the equilibrium concentration of conduction electrons from ionized donors is identical with the standard calculation in statistical mechanics of the thermal ionization of hydrogen atoms (*TP*, p. 369). If there are no acceptors present, the result in the low temperature limit $k_B T \ll E_d$ is

$$n \cong (n_0 N_d)^{1/2} \exp(-E_d/2k_B T) , \quad (53)$$

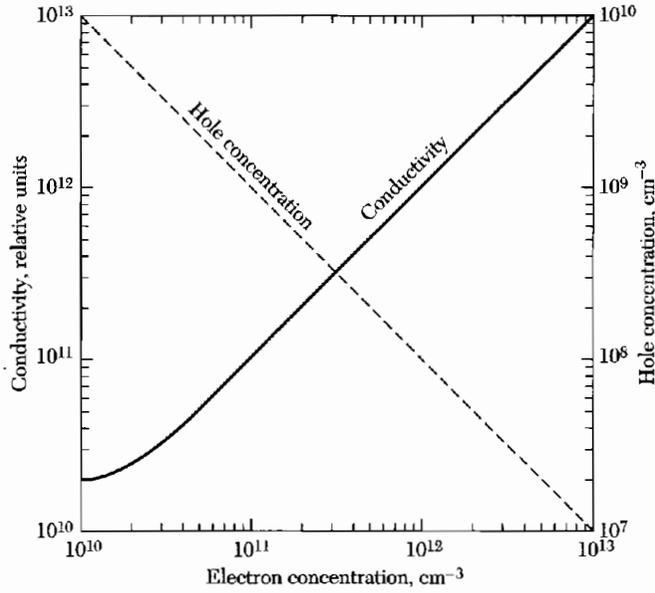


Figure 22 Electrical conductivity and hole concentration p calculated as a function of electron concentration n for a semiconductor at a temperature such that $np = 10^{20} \text{ cm}^{-6}$. The conductivity is symmetrical about $n = 10^{10} \text{ cm}^{-3}$. For $n > 10^{10}$, the specimen is n type; for $n < 10^{10}$, it is p type. We have taken $\mu_e = \mu_h$, for the mobilities.

with $n_0 \equiv 2(m_e k_B T / 2\pi\hbar^2)^{3/2}$; here N_d is the concentration of donors. To obtain (53) we apply the laws of chemical equilibria to the concentration ratio $[e][N_d^+]/[N_d]$, and then set $[N_d^+] = [e] = n$. Identical results hold for acceptors, under the assumption of no donor atoms.

If the donor and acceptor concentrations are comparable, affairs are complicated and the equations are solved by numerical methods. However, the law of mass action (43) requires the np product to be constant at a given temperature. An excess of donors will increase the electron concentration and decrease the hole concentration; the sum $n + p$ will increase. The conductivity will increase as $n + p$ if the mobilities are equal, as in Fig. 22.

THERMOELECTRIC EFFECTS

Consider a semiconductor maintained at a constant temperature while an electric field drives through it an electric current density j_q . If the current is carried only by electrons, the charge flux is

$$j_q = n(-e)(-\mu_e)E = ne\mu_e E, \quad (54)$$

where μ_e is the electron mobility. The average energy transported by an electron is referred to the Fermi level μ ,

$$(E_c - \mu) + \frac{3}{2} k_B T,$$

where E_c is the energy at the conduction band edge. We refer the energy to the Fermi level because different conductors in contact have the same Fermi level. The energy flux that accompanies the charge flux is

$$j_U = n(E_c - \mu + \frac{3}{2}k_B T)(-e\mu_e)E \quad (55)$$

The **Peltier coefficient** Π is defined by $j_U = \Pi j_q$; or the energy carried per unit charge. For electrons,

$$\Pi_e = -(E_c - \mu + \frac{3}{2}k_B T)/e \quad (56)$$

and is negative because the energy flux is opposite to the charge flux. For holes

$$j_q = pe\mu_h E \quad ; \quad j_U = p(\mu - E_v + \frac{3}{2}k_B T)\mu_h E \quad (57)$$

where E_v is the energy at the valence band edge. Thus

$$\Pi_h = (\mu - E_v + \frac{3}{2}k_B T)/e \quad (58)$$

and is positive. Equations (56) and (58) are the result of our simple drift velocity theory; a treatment by the Boltzmann transport equation gives minor numerical differences.²

The **absolute thermoelectric power** Q is defined from the open circuit electric field created by a temperature gradient:

$$E = Q \text{ grad } T \quad (59)$$

The Peltier coefficient Π is related to the thermoelectric power Q by

$$\Pi = QT \quad (60)$$

This is the famous Kelvin relation of irreversible thermodynamics. A measurement of the sign of the voltage across a semiconductor specimen, one end of which is heated, is a rough and ready way to tell if the specimen is *n* type or *p* type (Fig. 23).

SEMIMETALS

In semimetals the conduction band edge is very slightly lower in energy than the valence band edge. A small overlap in energy of the conduction and valence bands leads to small concentration of holes in the valence band and of electrons in the conduction band (Table 7). Three of the semimetals, arsenic, antimony, and bismuth, are in group V of the periodic table.

Their atoms associate in pairs in the crystal lattice, with two ions and ten valence electrons per primitive cell. The even number of valence electrons

²A simple discussion of Boltzmann transport theory is given in Appendix F.

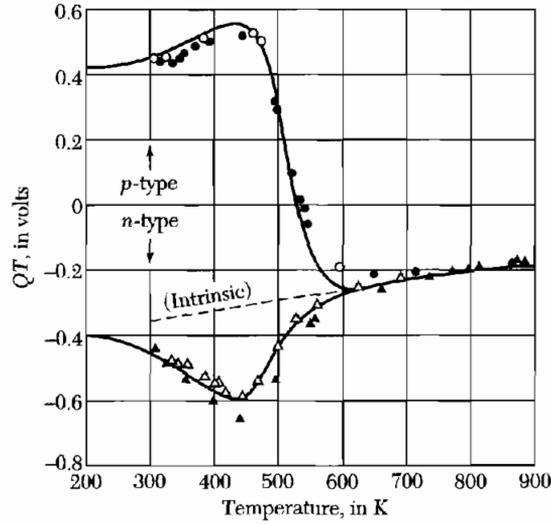


Figure 23 Peltier coefficient of *n* and *p* silicon as a function of temperature. Above 600 K the specimens act as intrinsic semiconductors. The curves are calculated and the points are experimental. (After T. H. Geballe and G. W. Hull.)

Table 7 Electron and hole concentrations in semimetals

Semimetal	n_n , in cm^{-3}	n_h , in cm^{-3}
Arsenic	$(2.12 \pm 0.01) \times 10^{20}$	$(2.12 \pm 0.01) \times 10^{20}$
Antimony	$(5.54 \pm 0.05) \times 10^{19}$	$(5.49 \pm 0.03) \times 10^{19}$
Bismuth	2.88×10^{17}	3.00×10^{17}
Graphite	2.72×10^{18}	2.04×10^{18}

could allow these elements to be insulators. Like semiconductors, the semimetals may be doped with suitable impurities to vary the relative numbers of holes and electrons. Their concentrations may also be varied with pressure, for the band edge overlap varies with pressure.

SUPERLATTICES

Consider a multilayer crystal of alternating thin layers of different compositions. Coherent layers on a nanometer thickness scale may be deposited by molecular-beam epitaxy or metal-organic vapor deposition, thus building up a superperiodic structure on a large scale. Systems of alternate layers of GaAs and GaAlAs have been studied to 50 periods or more, with lattice spacing A of perhaps 5 nm (50 Å). A superperiodic crystal potential arises from the superperiodic structure and acts on the conduction electrons and holes to create new (small) Brillouin zones and mini energy bands superposed on the band structures of the

constituent layers. Here we treat the motion of an electron in a superlattice in an applied electric field.

Bloch Oscillator

Consider a collisionless electron in a periodic lattice in one dimension, with motion normal to the planes of the superlattice. The equation of motion in a constant electric field parallel to k is $\hbar dk/dt = -eE$ or, for motion across a Brillouin zone with reciprocal lattice vector $G = 2\pi/A$, we have $\hbar G = \hbar 2\pi/A = eET$, where T is the period of the motion. The **Bloch frequency** of the motion is $\omega_B = 2\pi/T = eEA/\hbar$. The electron accelerates from $k = 0$ toward the zone boundary; when it reaches $k = \pi/A$ it reappears (as by an Umklapp process) at the zone boundary at the identical point $-\pi/A$, using the argument of Chapter 2.

We consider the motion in a model system in real space. We suppose that the electron lies in a simple energy band of width ϵ_0 :

$$\epsilon = \epsilon_0(1 - \cos kA). \quad (61)$$

The velocity in k -space (momentum space) is

$$v = \hbar^{-1} d\epsilon/dk = (A\epsilon_0/\hbar) \sin kA, \quad (62)$$

and the position of the electron in real space, with the initial condition $z = 0$ at $t = 0$, is given by

$$\begin{aligned} z &= \int v dt = \int dk v(k)(dt/dk) = (A\epsilon_0/\hbar) \int dk (-\hbar/eE) \sin kA \\ &= (-\epsilon_0/eE)(\cos kA - 1) = (-\epsilon_0/eE)(\cos(-eEA t/\hbar) - 1). \end{aligned} \quad (63)$$

This confirms that the Bloch oscillation frequency in real space is $\omega_B = eEA/\hbar$. The motion in the periodic lattice is quite different from the motion in free space, for which the acceleration is constant.

Zener Tunneling

Thus far we have considered the effect of the electrostatic potential $-eEz$ (or $-eEnA$) on one energy band; the potential tilts the whole band. Higher bands will also be tilted similarly, creating the possibility of crossing between ladder levels of different bands. The interaction between different band levels at the same energy opens the possibility for an electron in one band at n to cross to another band at n' . This field-induced interband tunneling is an example of **Zener breakdown**, met most often at a single junction as in the Zener diode.

SUMMARY

- The motion of a wave packet centered at wavevector \mathbf{k} is described by $\mathbf{F} = \hbar d\mathbf{k}/dt$, where \mathbf{F} is the applied force. The motion in real space is obtained from the group velocity $\mathbf{v}_g = \hbar^{-1} \nabla_{\mathbf{k}} \epsilon(\mathbf{k})$.

- The smaller the energy gap, the smaller is the effective mass $|m^*|$ near the gap.
- A crystal with one hole has one empty electron state in an otherwise filled band. The properties of the hole are those of the $N - 1$ electrons in this band.
 - (a) If the electron is missing from the state of wavevector \mathbf{k}_e , then the wavevector of the hole is $\mathbf{k}_h = -\mathbf{k}_e$.
 - (b) The rate of change of \mathbf{k}_h in an applied field requires the assignment of a positive charge to the hole: $e_h = e = -e_e$.
 - (c) If \mathbf{v}_e is the velocity an electron would have in the state \mathbf{k}_e , then the velocity to be ascribed to the hole of wavevector $\mathbf{k}_h = -\mathbf{k}_e$ is $\mathbf{v}_h = \mathbf{v}_e$.
 - (d) The energy of the hole referred to zero for a filled band is positive and is $\epsilon_h(\mathbf{k}_h) = -\epsilon(\mathbf{k}_e)$.
 - (e) The effective mass of a hole is opposite to the effective mass of an electron at the same point on the energy band: $m_h = -m_e$.

Problems

1. **Impurity orbits.** Indium antimonide has $E_g = 0.23$ eV; dielectric constant $\epsilon = 18$; electron effective mass $m_e = 0.015 m$. Calculate (a) the donor ionization energy; (b) the radius of the ground state orbit. (c) At what minimum donor concentration will appreciable overlap effects between the orbits of adjacent impurity atoms occur? This overlap tends to produce an impurity band—a band of energy levels which permit conductivity presumably by a hopping mechanism in which electrons move from one impurity site to a neighboring ionized impurity site.
2. **Ionization of donors.** In a particular semiconductor there are 10^{13} donors/cm³ with an ionization energy E_d of 1 meV and an effective mass 0.01 m . (a) Estimate the concentration of conduction electrons at 4 K. (b) What is the value of the Hall coefficient? Assume no acceptor atoms are present and that $E_g \gg k_B T$.
3. **Hall effect with two carrier types.** Assuming concentration n, p ; relaxation times τ_e, τ_h ; and masses m_e, m_h , show that the Hall coefficient in the drift velocity approximation is

$$(CGS) \quad R_H = \frac{1}{ec} \cdot \frac{p - nb^2}{(p + nb)^2},$$

where $b = \mu_e/\mu_h$ is the mobility ratio. In the derivation neglect terms of order B^2 . In SI we drop the c . Hint: In the presence of a longitudinal electric field, find the transverse electric field such that the transverse current vanishes. The algebra may seem tedious, but the result is worth the trouble. Use (6.64), but for two carrier types; neglect $(\omega_c \tau)^2$ in comparison with $\omega_c \tau$.

- 4. Cyclotron resonance for a spheroidal energy surface.** Consider the energy surface

$$\epsilon(\mathbf{k}) = \hbar^2 \left(\frac{k_x^2 + k_y^2}{2m_t} + \frac{k_z^2}{2m_l} \right),$$

where m_t is the transverse mass parameter and m_l is the longitudinal mass parameter. A surface on which $\epsilon(\mathbf{k})$ is constant will be a spheroid. Use the equation of motion (6), with $\mathbf{v} = \hbar^{-1} \nabla_{\mathbf{k}} \epsilon$, to show that $\omega_c = eB/(m_t m_l)^{1/2} c$ when the static magnetic field B lies in the xy plane. This result agrees with (34) when $\theta = \pi/2$. The result is in CGS; to obtain SI, omit the c .

- 5. Magnetoresistance with two carrier types.** Problem 6.9 shows that in the drift velocity approximation the motion of charge carriers in electric and magnetic fields does not lead to transverse magnetoresistance. The result is different with two carrier types. Consider a conductor with a concentration n of electrons of effective mass m_e and relaxation time τ_e ; and a concentration p of holes of effective mass m_h and relaxation time τ_h . Treat the limit of very strong magnetic fields, $\omega_c \tau \gg 1$. (a) Show in this limit that $\sigma_{yx} = (n - p)ec/B$. (b) Show that the Hall field is given by, with $Q \equiv \omega_c \tau$,

$$E_y = -(n - p) \left(\frac{n}{Q_e} + \frac{p}{Q_h} \right)^{-1} E_x,$$

which vanishes if $n = p$. (c) Show that the effective conductivity in the x direction is

$$\sigma_{\text{eff}} = \frac{ec}{B} \left[\left(\frac{n}{Q_e} + \frac{p}{Q_h} \right) + (n - p)^2 \left(\frac{n}{Q_e} + \frac{p}{Q_h} \right)^{-1} \right].$$

If $n = p$, $\sigma \propto B^{-2}$. If $n \neq p$, σ saturates in strong fields; that is, it approaches a limit independent of B as $B \rightarrow \infty$.